知识管理论坛 ISSN 2095-5472 CN11-6036/C
Knowledge Management Forum

E-mail: kmf@mail.las.ac.cn
http://www.kmf.ac.cn

【学术探索】

基于知识实体的突发公共卫生事件数据平台构建研究

冯鑫 [1] 李雪 [2] 闫月 [2] 李佳培 [1] 刘梦瑶 [1] 吴晔 [3,4]

[1] 河北地质大学管理学院 石家庄 050031
[2] 河北地质大学华信学院 石家庄 050031
[3] 北京师范大学计算传播学研究中心 珠海 519087
[4] 北京师范大学新闻传播学院 北京 100875

摘要：[**目的 / 意义**] 突发公共卫生事件严重危及全球社会公众的生命健康。实现信息资源对突发事件的智能存储、查询和知识组织与输出等，对应急领域科研数据集成共享与领域知识管理具有重要的参考意义。[**方法 / 过程**] 选取 CNKI 和 WOS 为检索平台，对突发公共卫生事件进行检索，爬取有效数据，利用文献计量软件对抽取的知识数据进行加工、融合，并对高频词汇、文献作者和机构进行聚类可视化分析，将知识实体、属性及关系存储于 Neo4j 图数据库中，建立知识实体之间的复杂联系，进而搭建突发公共卫生事件数据平台模型。[**结果 / 结论**] 应急事件智慧平台是一个以信息资源为基础，对来自不同领域的信息进行处理，通过内部实体的筛选与融合构建的面向大众的知识数据共享资源平台，该平台可以上下交互实现信息资源流动，切实发挥信息资源对应急管理工作的作用。

关键词：突发公共卫生事件 应急服务 知识实体 2019 新型冠状病毒
分类号：G203
DOI：10.13266/j.issn.2095-5472.2020.016

引用格式：冯鑫 , 李雪 , 闫月 , 等 . 基于知识实体的突发公共卫生事件数据平台构建研究 [J/OL]. 知识管理论坛 , 2020, 5(3): 175-190[引用日期]. http://www.kmf.ac.cn/p/211/.

基金项目：本文系教育部人文社会科学研究青年基金项目"基于复杂网络及数据挖掘的智慧教育规划研究"（项目编号：16YJC630022）、2018 年度河北省社会科学发展研究课题"基于阅读行为动力学的移动媒体舆情管控"（项目编号：201802120102）和石家庄市科学技术研究与发展计划项目"基于知识图谱的科普大数据平台规划研究"（项目编号：191790115A）研究成果之一。

作者简介：冯鑫（ORCID：0000-0002-0140-8003），副教授，博士，硕士生导师；李雪（ORCID：0000-0001-9837-4788），本科生；闫月（ORCID：0000-0003-1726-315X），本科生；李佳培（ORCID：0000-0002-8720-6022），讲师，博士，通讯作者，Email：jiapeil@gmail.com；刘梦瑶（ORCID：0000-0003-4673-9954），本科生；吴晔（ORCID：0000-0001-9038-2900），教授，博士。

收稿日期：2020-06-04 发表日期：2020-06-26 本文责任编辑：刘远颖



知识管理论坛
2020 年 第 3 期(总第 27 期)

1 引言

突发公共卫生事件，是指突然发生，造成或者可能造成社会公众健康严重损害的重大传染病疫情、群体性不明原因疾病、重大食物和职业中毒以及其他严重影响公众健康的事件，如当前热点事件——新型冠状病毒肺炎（以下简称“新冠”）。2020 年 1 月 30 日晚，世界卫生组织（以下简称“世卫组织”）总干事谭德赛在日内瓦举行新闻发布会，声明新型冠状病毒感染的肺炎疫情已构成国际关注的突发公共卫生事件。新冠的爆发已轰动全国乃至全世界，疫情的肆意蔓延以及进一步恶化将造成严重的经济冲击和社会影响，威胁着世界人民的生命健康。

世界卫生组织在发布会上表示，在习近平主席亲自指挥、亲自部署疫情防控的组织下，中方始终坚持公开透明原则，及时发布信息，快速识别病毒并分享基因序列，采取果断有力的措施控制疫情传播。这些不仅是对本国人民生命健康的高度负责，更是对全球疾病防控的大力支持。此外，在本次疫情中，中国采取的大量防控措施远远超出国际应对突发事件的相关要求，为各国防疫工作设立了新标杆。就目前来看，中国的疫情已得到了有效控制，但是国际的蔓延是对各国家的又一个挑战。世卫组织每日疫情报告显示，截至欧洲中部时间 2020 年 6 月 12 日 10 时（北京时间 16 时），中国以外新冠确诊病例较前一日增加 136 565 例，达到 7 325 851 例；中国以外死亡病例较前一日增加 49 25 例，达到 413 649 例。全球范围内，新冠确诊病例较前一日增加 136 572 例，达到 7 410 510 例；死亡病例较前一日增加 4 925 例，达到 418 294 例。面对突发公共卫生事件的严峻形势，必须采取预防为主、防治结合的举措，加强联防联控，制定完善的应对计划。习总书记在改革完善重大疫情防控救治体系中提到，国家鼓励运用大数据、人工智能、云计算等数字技术，在疫情监测分析、病毒溯源、防控救治、资源调配等方面更好地发挥其支撑作用 [1]。

因此，有必要缜密调研国内外突发公共卫生事件的研究现状，抽取复杂抽象的多维数据，以实体、关系和属性的方式将专业知识实体关联化和可视化，并基于精准全面知识实体来构建应急智慧数据平台。通过搭建平台，从低密度的数据中提取有价值的信息进行及时共享和智慧分享，使得相关信息的获取变得方便迅速且高效，最终实现智能搜索和智能推荐帮助多类型用户快速精准地收集信息，利用信息资源制定应急防控措施，对于减少突发公共卫生事件造成的损失具有重要现实意义。

2 相关研究

科研学术机构可以通过其自身职能和特点支持和响应公共卫生对策，做出信息管理和绩效评估。A. L. Dunlop 等通过回顾 2001 年 9 月 11 日至 2009 年 2 月 1 日的文献和互联网报告，提出了学术机构作为社会灾害响应的一部分，提供的服务和资源可以减少事件的不良后果、降低其发生的可能性，从而有助于提高社区的抗灾能力 [2]。N. A. Vielot 等对于合并公共卫生防范和应急管理的作用是否可以完善应急计划和提高响应的效率和效力这一问题，在美国北卡罗来纳州的 6 个县进行采访，讨论产生的机遇和挑战 [3]。J. Hu 等研究建立了具有公共卫生管理职能的农村应急管理常设机构，完善了应急管理机制的全过程，加大了对农村公共卫生的投资力度，构建了具有充足应急资源储备的体系 [4]。S. Zhong 等基于传统的云计算在应急服务中存在延迟等缺陷，提出了构建一种基于边缘计算模型的紧急救援体系结构，并在此基础上提供一个资源调度模型，从而使其具有良好的调度时间和较低的成本 [5]。

国内学者也对公共突发事件的资源建设进行了深入分析。在突发事件应急情报资源保障方面，柯丹倩对国内外突发事件驱动的应急决策情报现状进行了综述分析，为后续研究做出铺垫 [6]。张永领等将情景分析引入到评价模型，对应急资源的保障能力进行全方位评价，从而

知识管理论坛，2020(3):175–190
DOI：10.13266/j.issn.2095–5472.2020.016

使评价结果更接近实际需求[7]。在实现应急信息资源共享、完善应急管理方面，郭陆生等通过构建目录，为应急信息资源提供统一的标准体系，实现了信息资源与应急服务的衔接，为应急信息资源跨部门共享与相互服务提供了便利[8]。屈腾佼等对中国 2003-2018 年重特大突发公共卫生事件案例和相关卫生应急管理文献进行归纳分析，阐述了中国卫生应急管理现状[9]。在应急决策和信息融合方面，李品等实现了智库活动与情报研究内部的融合，构建了支持智库活动的情报流程并协助智库产品产生的情报体系[10]。操玉杰等对应急决策信息需求与大数据环境下可获得的应急信息进行对比，构建以应急数据模型为桥梁、面向应急决策的信息融合服务整体框架；逐层剖析数据层、语义层和服务层应急信息融合中的具体问题，提出了相应的信息融合实现路径。该研究结论有助于指导突发事件大数据融合系统构建，也为面向决策的信息分析提供了实践参考依据[11]。

综上所述，现有的研究成果充分说明，在信息化时代应急信息资源对应对突发公共卫生事件有着不可或缺的作用，而目前还没有相关学者对应急资源进行知识实体的构建。因此，本文的研究将从中国知网（CNKI）和 WOS（Web of Science）中提取知识实体，基于知识实体之间的关系建立知识实体图，然后在数据库底层进行逻辑推理，把突发公共卫生事件相关的知识体系系统化地展示给用户，通过知识实体图辅助科普领域大数据的分析与决策，使应急资源的组织更加高效，从而提供更加精准的服务。

3 研究对象与数据来源

本文对国内中国知网和国外 WOS 进行主题检索。首先，在中国知网学术期刊总库检索中，采用高级检索的方式，选择的检索字段为“关键词”，检索词为“突发公共卫生事件”，限定论文为 SCI、EI 核心期刊，采集时间段为 2010 年到 2020 年，剔除无关文献后得到检索结果为 416 篇。其次，在 Web of Science 全文数据库检索中，选择 Web of Science 核心合集，检索文章的主题为“public health emergencies”，限定文献类型为 Article OR Review，索引为 SCI-EXPANDED、SSCI、A&HCI，语种为英语，采集时间段为 2010 年到 2020 年，剔除无关文献后得到的检索结果为 949 篇。获取相关数据后，采用可视化的方法，从文献数量、引文数量、作者和机构对比等方面分析，通过科学计量指标，对突发公共卫生事件进行量化分析。

4 知识实体的提取

4.1 知识实体提取方法

科研论文中的关键词表征了一项科研成果的核心研究内容，是检索知识的重要依据。本研究利用文献计量可视化工具 Citespace 对中外文相关主题数据集进行计量统计，析出核心关键词作为知识节点之一，析出核心专家与研究机构作为节点之一。以关键词“突发公共卫生事件”为主要对象，搭建知识实体的主要框架；对文档内容进行分析，以自动提取的关键词为对象，建立关联度强的知识实体；进行用户关联分析，挖掘潜在的知识关联，并将普遍关联的知识实体应用于科普平台搭建，对科普用户进行科普知识个性化推荐。以构建知识实体的应急资源平台为出发点，从科普知识实体抽取，到实体关系构建，知识实体可视化，最终实现信息共享、智能查询、智能推荐等功能，提高科普信息的利用率。

本研究利用关键词共现聚类形成网络，中文数据集聚类如图 1 所示，共计 8 个聚类单元，分别为：突发公共卫生事件、学生保健服务、传染病、重大突发公共卫生事件、大数据、公共危机预警、delphi 法、水痘。外文数据集聚类如图 2 所示，共计 7 个聚类，分别为 resilience、ebola、synthetic cannabinoids、radiological、emergency preparedness、quarantine、practice guidelines。






图 1 中文突发公共卫生事件数据集聚类

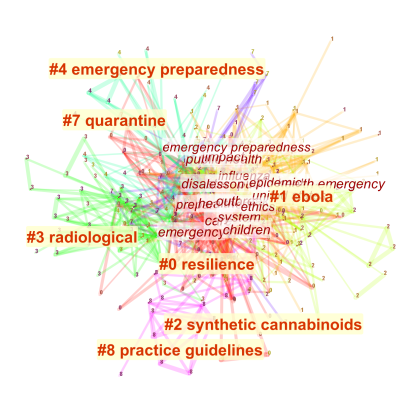


图 2 外文突发公共卫生事件数据集聚类

4.2 机构发文量分析

统计各个机构的发文量，得到主要研究机构名单。在 Citespace 中选择节点 Institution，每个节点的时间切片选择 Top50 的点节点标签分别分析国内与国外的机构发文量，得到发文量统计表和机构研究合作图。

表 1 是 2010 年到 2020 年国内突发公共卫生事件机构的发文量排名前十的统计表，排名前三的机构分别是武汉大学信息管理学院（10 篇）、江苏省疾病预防控制中心（7 篇）、哈尔滨医科大学卫生管理学院社会医学教研室（5 篇），这 3 所机构的发文量都超过了 5

知识管理论坛，2020(3):175–190
DOI：10.13266/j.issn.2095–5472.2020.016

篇。表 2 研究的是从 2010 年到 2020 年期间国外突发公共卫生事件机构的发文量排名。从统计表的年份看出，国内外突发公共卫生事件机构的发文量大部分处于这 10 年间的前 5 年，说明各大机构对突发公共卫生事件领域涉及时间都很早。

表 1　国内突发公共卫生事件机构发文量前十统计表

排序	机构名称	发文量/篇	年份区间
1	武汉大学信息管理学院	10	2017-2020
2	江苏省疾病预防控制中心	7	2010-2020
3	哈尔滨医科大学卫生管理学院社会医学教研室	5	2013-2020
4	中国疾病预防控制中心环境与健康相关产品安全所	4	2010-2020
5	杭州市疾病预防控制中心	3	2015-2020
6	湖北省疾病预防控制中心	3	2010-2020
7	清华大学应急管理研究基地	3	2020
8	哈尔滨医科大学卫生管理学院	3	2013-2020
9	武汉大学信息资源研究中心	3	2020
10	北京市通州区疾病预防控制中心	3	2010-2020

图 3 是国内突发公共卫生事件的机构合作图，本图是利用 Citespace 对国内的机构合作进行可视化分析得到的。根据节点半径大小得出，武汉大学信息管理学院是这些机构中发文量最多的机构，其次是江苏省疾病预防控制中心，哈尔滨医科大学卫生管理学院社会医学教研室紧随其后。但之间的合作却寥寥无几。相比之下武汉大学信息资源研究中心与排名第一的武汉大学信息管理学院的合作更多一些。从这些机构发文量的分布来看，发文机构遍布全国各地，说明我国对突发公共卫生事件的研究机构地理分布比较均匀，可以做到按照本区域公共卫生的特点进行区域研究，但国内尚未形成比较深入和密集的跨区域跨机构研究局面。



图 3　突发公共卫生事件国内机构合作图

近几年随着突发公共卫生事件逐渐被重视，国外的一些机构也对此领域研究下足了功夫。表 2 是国外研究机构突发事件发文量排名前十的统计表，是基于 Web of Science 近 10 年的机构发文量统计。其中 Ctr Dis Control & Prevent（美国疾病预防控制中心）的总发文量达到了 88 篇，WHO（世界卫生组织）39 篇、Harvard Univ（哈佛大学）34 篇、Minist Hlth（沙特阿拉伯卫生部）24 篇、CDC（疾病预防控制中心）23 篇。

表 2 国外突发公共卫生事件机构发文量前十统计表

排序	机构名称	发文量	年份区间
1	Ctr Dis Control & Prevent	88	2010-2020
2	WHO	39	2010-2020
3	Harvard Univ	34	2011-2020
4	Minist Hlth	24	2011-2020
5	CDC	23	2010-2020
6	Univ Pittsburgh	20	2010-2020
7	Emory Univ	19	2012-2020
8	Johns Hopkins Bloomberg Sch Publ Hlth	19	2012-2020
9	Univ Toronto	17	2010-2020
10	Univ Penn	14	2015-2020

图 4 是利用 Citespace 对 Web of science 核心合集中突发公共卫生事件文献可视化，得到的国外机构合作的可视化分析图，线的粗细表示机构之间的联系，节点的大小表示机构发文频次的高低。图中线条节点都紧密联系在一起，说明国外的各大机构之间的合作力度较强。



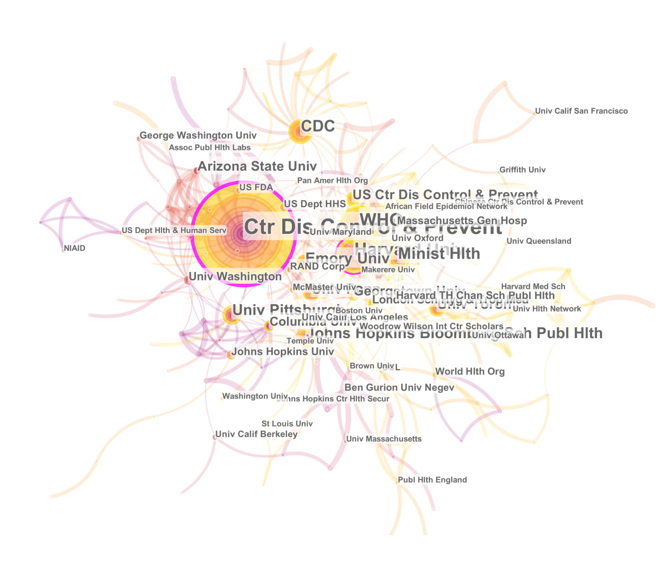


图 4 突发公共卫生事件国外机构合作图

知识管理论坛，2020(3):175–190
DOI：10.13266/j.issn.2095–5472.2020.016

4.3 作者发文量分析

在 Citespace 中选择 Author，时间切片选择 Top50 的点，选择 Thresholding 为（2，1，20），设置节点参数发文量为至少 3 篇。分析 2010 年 -2020 年的作者发文量，设置年份间隔为 1，得到图 5；分析 2010 年 -2020 年的作者发文量，将设置年份间隔为 1，得到图 6。

表 3 是经 Citespace 统计分析得出发文量排名前十的作者统计表，可以看出吴群红在这 10 年的发文总量达到 10 篇，安璐 7 篇，刘静 5 篇，郝艳华 5 篇。说明这些科研人员在突发公共卫生事件的研究领域做出了较大贡献。

用 Citespace 对作者进行可视化分析后得出国内作者合作分析图（见图 5），其中颜色随发表时间由深变浅。本图显示发表 2 篇文献以上的专家名称，节点半径越大，名字字体越大，代表作者发表的文献数量越多。连接线越密集说明作者之间的合作越紧密。在图中，黄色代表浅色，结合表 3 可以得知安璐是 2016 年开始涉及突发事件领域，并且发文成果可观，对此领域做出了贡献。

表 3 国内突发公共卫生事件发文量前十作者统计表

排序	作者	发文量/篇	年份区间
1	吴群红	10	2013-2020
2	安璐	7	2016-2020
3	刘静	5	2013-2020
4	郝艳华	5	2013-2020
5	宁宁	4	2013-2020
6	周志衡	4	2012-2020
7	胥娇	4	2013-2020
8	黄春萍	3	2015-2020
9	徐珏	3	2015-2020
10	刘敏	3	2011-2020

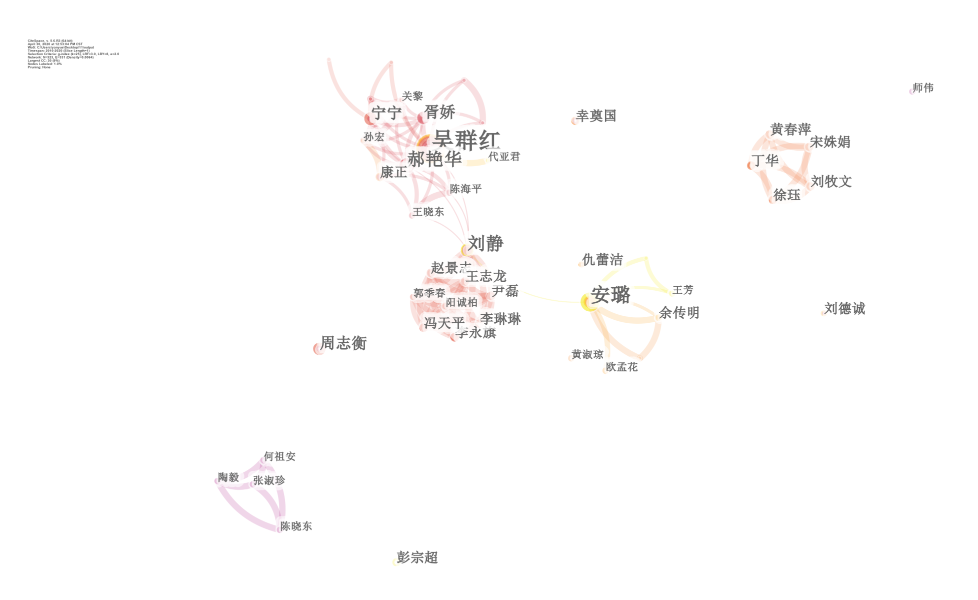


图 5 突发公共卫生事件国内作者合作分析

表 4 是 WOS 中 2010-2020 年期间排名前十的作者统计。从表中可以看出发文最多的是 FREDERICK M（7 篇），其研究方向是预防医学与卫生学。其次是 ELENA SAVOIA（5 篇），主要从事预防医学、公共卫生和定量方法研究；DANIEL J BARNETT（5 篇）主要专业是急诊医学；LAINIE RUTKOW（5 篇）主要采用法律、定性和定量的研究方法，对公共卫生法律和政策进行研究。

图 6 显示了 WOS 中作者的分布。如图

所示，节点越大表示作者发文量越多，节点年轮的颜色变化反映了作者活跃的不同时期。从图中可以看出，这些作者的联系并不紧密，发文量前三名的作者几乎没有合作关系，都在以独自的思维方式去进行研究。

表 4 国外突发公共卫生事件发文量前十作者统计表

排序	作者	发文量	年份区间
1	FREDERICK M	7	2011-2020
2	ELENA SAVOIA	5	2014-2020
3	DANIEL J BARNETT	5	2010-2020
4	LAINIE RUTKOW	5	2014-2020
5	KATHERINE SEIB	4	2012-2020
6	SAAD B OMER	4	2012-2020
7	NIRANJAN KISSOON	4	2011-2020
8	JAMES G	4	2011-2020
9	ZHANAT CARR	3	2010-2020
10	JON S VERNICK	3	2014-2020

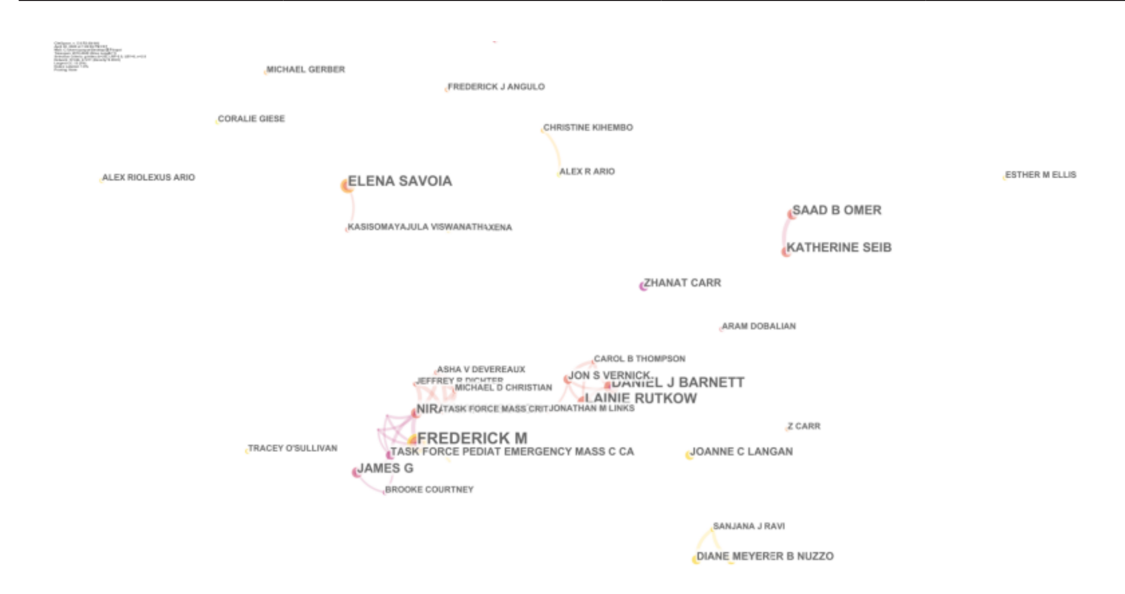


图 6 突发公共卫生事件国外作者合作分析

4.4 突发公共卫生事件知识实体构建

4.4.1 知识实体图相关概念

知识实体图，又称为知识域可视化或知识领域映射地图，是显示知识发展进程与结构关系的一系列各种不同的图形。用可视化技术描述知识资源及其载体，挖掘、分析、构建、绘制并显示知识及其相互联系。其本质就是一种大规模的语义网络，富含实体、概念及关系，是大规模数据知识表示的方法之一。

知识实体可以存储总结不断迭代的知识、

知识管理论坛，2020(3):175–190
DOI：10.13266/j.issn.2095–5472.2020.016

规则、元数据等多类型的数据或者融合结构化与非结构化的数据，可以利用它来分析关系复杂的多维数据。知识实体图主要有两种存储方式：一种是基于 RDF[12] 的存储，一种是基于图数据库的存储。RDF 以三元组方式来存储数据而且不包含属性信息，但图数据库一般以属性图为基本的表示形式，实体和关系可以包含属性，这样更容易表达现实的业务场景，它们之间的主要区别就是 RDF 中的一个重要设计原则是数据的易发布及共享，而图数据库则把重点放在了高效的图查询和搜索上。图 7 和图 8 就是基于 Neo4j[13] 图数据库做出的知识实体图。

Neo4j 是目前使用率最高的图数据库，以描述实体的属性与实体间的关系为主，用图的形式存储。它拥有自己的数据库存储，支持 320 亿的关系节点和 640 亿的属性。Neo4j 拥有活跃的社区，查询效率高，但缺点为分布式存储实现代价高，更新速度慢。Neo4j 也是一种非关系型数据库，它用“节点”和“关系”来存储数据与数据间的关联，由于不需要固定的 schema，可以随项目发展扩展节点上的属性。

本研究以 CNKI 和 WOS 提取的数据集中的主要关键词、研究者和机构聚类名称，作为实体节点名称。全部融合考虑以上知识节点后，将突发公共卫生事件知识实体分为 5 部分，即突发公共卫生级别、防控措施、传染病、专家和机构。然后定义实体与实体、实体与属性之间的关系，用 CREATE 语句创建节点、属性和关系。最后，我们将所有命名的实体及实体关系导入 Neo4j 图数据库，得到图 7 突发公共卫生事件知识实体图和图 8 新冠在中西医研究方面知识融合实体图两个重要知识实体图。

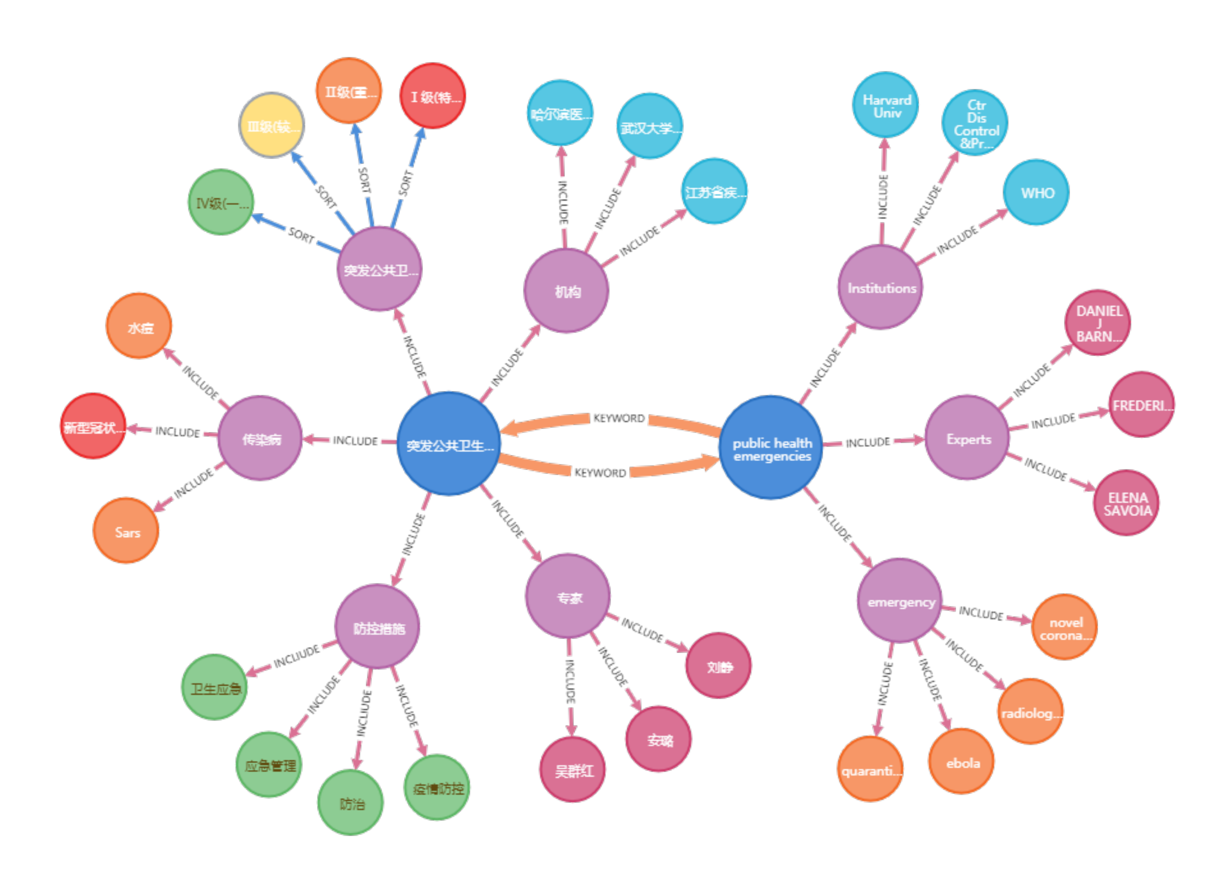


图 7 突发公共卫生事件知识实体图



知识管理论坛
2020 年 第 3 期(总第 27 期)

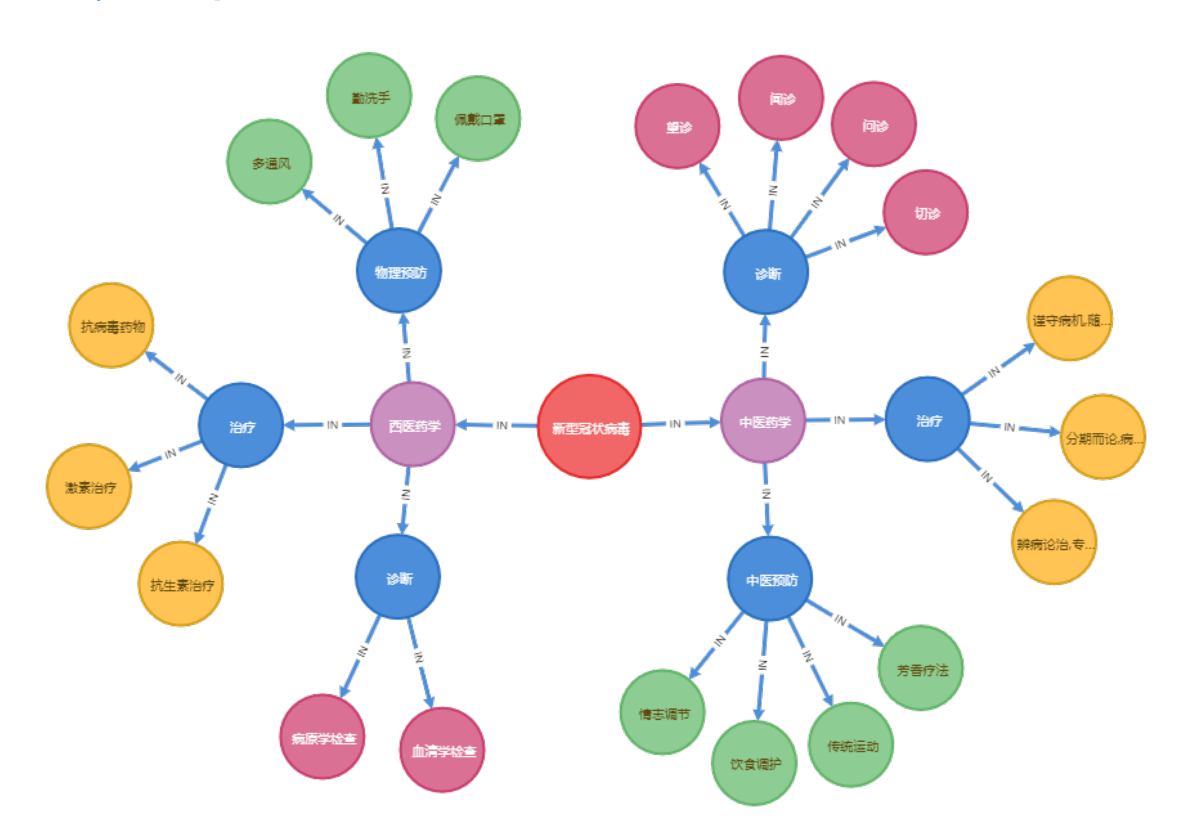


图 8 新冠在中西医研究方面知识融合实体图

4.4.2 突发公共卫生事件知识实体的构建

图 7 是突发公共卫生事件的知识实体图，由节点和边构成的，其中圆圈代表一个节点，即知识实体中的实体，实体之间的边代表属性与节点的关系，关系可以用箭头来表示方向，两端对应开始节点和结束节点。在节点上加一个或多个标签表示实体的分类，和一个键值对集合来表示实体关系属性外的其他属性，这时关系也可以附带额外属性。我们一般都是将实体的 property-value（属性值）表示在图中，由于属性和节点分开储存，其属性可以用 MARCH 命令和 WHERE 子句来迅速实现查询，部分节点属性如表 5 所示。

突发公共卫生事件知识实体图是根据图 1 和图 2 的关键词聚类分析、图 3 和图 5 的作者合作分析以及图 4 和图 6 的机构合作分析来创建的，它的两个一级知识实体分别为“突发公共卫生事件”与“Public health emergencies”，用最大的圈来展示且设置为蓝色，中英两个一级实体的设定是为后续知识实体在中外文两方面同时展开增长做好准备工作。根据上文的关键词聚类可视化分析，本文把突发公共卫生事件分为五大部分：突发公共卫生事件级别、防控措施、传染病、专家和机构，将这 5 个实体与“突发公共卫生事件”的实体的关系属性设为“INCLUDE”，即是“属于”上一级的下位类。《国家突发公共卫生事件应急预案》第 1.3 条第 2 项规定，根据突发公共卫生事件性质、危害程度、涉及范围，突发公共卫生事件可划分为特别重大（Ⅰ级）、重大（Ⅱ级）、较大（Ⅲ级）和一般（Ⅳ级）四级[14]，“特别重大”用红色表示、“重大”用橙色表示、“较大”用黄色表示、“一般”用绿色表示，设置这 4 个实体节点与“突发公共卫生事件级别”实体节点的关系属性为“SORT（分类）”。在防控措施方面，我国积极做到了“应

知识管理论坛，2020(3):175–190
DOI：10.13266/j.issn.2095–5472.2020.016

急管理”“卫生应急”“疫情防控”和“防治”，这 4 个实体的聚类程度和实践意义都比较强，用绿色的节点表示。突发公共卫生事件最严重的事件就是“传染病”，其中对人类造成巨大危害的有“新冠”“Sars”“水痘”等。今年爆发的“新冠”使我国多个城市做出了一级响应，所以用红色圈重点标识。根据上文作者发文统计表和机构发文统计表，将专家和机构发文量的前三名放进知识实体库中分别用紫红色和蓝色表示。

表 5 突发公共卫生事件知识实体属性表

ID	P1_name	P1_value	P2_name	P2_value	P3_name	P3_value
4003	Keyword	突发公共卫生事件	Search range	416	From	中国知网
4004	Keyword	Public Health Emergencies	Search range	964	From	Web of science
4005	Dept	emergency	Branch Number	4	nodeType	String
4010	Dept	Experts	Branch Number	3	nodeType	String
4014	Dept	Institutions	Branch Number	3	nodeType	String
4018	Dept	传染病	Branch Number	3	nodeType	String
4019	Dept	防控措施	Branch Number	4	nodeType	String
4020	Dept	突发公共卫生事件级别	Branch Number	4	nodeType	String
4021	Dept	专家	Branch Number	3	nodeType	String
4022	Dept	机构	Branch Number	3	nodeType	String
4034	Name	吴群红	Total literatures	360	Organization	哈尔滨医科大学
4035	Name	安璐	Total literatures	67	Organization	武汉大学
4036	Name	刘静	Total literatures	37	Organization	郑州大学

对于国外部分的知识实体建设，本文主要根据聚类结果将“Public health emergencies”事件分为三部分，第一部分为“Emergency”，包括 ebola、radiological、novel coronavirus 和 quarantine 等，第二部分和第三部分分别为“Expers”和“Institutions”，根据国外作者发文统计表将作者 FREDERICK M、ELENA SAVOIA 和 DANIEL J BARNETT 等和国外机构发文机构加入知识实体。

4.4.3 突发公共卫生事件的节点属性

表 5 是突发公共卫生事件知识实体图部分节点属性的展列，属性是描述节点或边的性质，可以用键值对（key-value）[15] 来表示。Neo4j 在创建属性时分为两部分 Property-name（属性名称）和 Property-value（属性的值）两部分，同一属性名可对应多个属性值。突发公共卫生事件知识实体属性表（见表 5），表头是由节点的 ID、3 个属性名称和其对应的属性值构成的。其中，知识节点的 ID 是在创建节点时自动形成，且 ID 是唯一的。属性表的第一列的元素是每个节点的 ID, 第二、四、六列是指对每个节点设置的属性名称（P1_name、P2_name、P3_name），第三、五、七列分别是其前一列的属性值（P1_value、P2_value、P3_value）。一般情况下，节点有多个属性时，系统会自动选择 name 属性的属性值在图上显示，如：Person:person{age:20,sex:”woman”,name:”Lisa”}，最终在图中显示的是“Lisa”。将表 5 第二列的属性名称对应的属性值显示在图 7 中。除图中显示的属性值，属性表将所有含有下位类的节点及部分子节点的所有属性都罗列出来了，如 ID 为 4003 的属性名 keyword 对应的属性值是突发



公共卫生事件，属性名 search range 检索量对应的属性值为 1278，属性名 from 来源对应属性值为中国知网。

4.4.4 知识实体查询

Neo4j 图数据库通过 Cypher[16] 语言实现对图的查询操作。在图形数据库 Neo4j 中，图遍历查询的速度是一个独立于遍历数据量大小的常数，仅当需要访问某节点或关系时，图形数据库 Neo4j 才会对其进行遍历并返回节点。当使用 Cypher 语句 MATCH 匹配正则表达式时，Cypher 语句是一个描述性的图查询语言，会自动调用内置的遍历搜索算法遍历访问，不需要人为指定遍历搜索方式，即可遍历得到整张图。

图 8 是对图 7 使用 Cypher 查询语句得到的网络图，是参考《新型冠状病毒肺炎文献整理及研究概述》[17] 和《基于中医经典理论探讨新型冠状病毒肺炎的防治》[18]，对新冠展开的中西医药学的诊断、治疗和预防研究的概括总结。其查询语句是：

```
MATCH(Sara:sara)
MATCH(Sara:sara)-[r1:IN]->(n1)-[r2:IN]->(n2)-[r3:IN]->(n3)
RETURN Sara,n1,r1,n2,r2,n3,r3
```

MATCH 和 RETURN 是 Cypher 的关键字；Sara 是变量用来保存节点；sara 是标签；r1，r2，r3 代表节点之间的关系；n1，n2，n3 是指与上一位节点关系为“IN”的所有节点。当实行第一段语句时，查询得到一个名为“新型冠状病毒”的红色节点；随后执行到“n1”时，经数据库查询显示“中医药学”和“西医药学”两个节点（用紫色表示）；执行到“n2”时，查询得到“中医药学”和“西医药学”两个节点的所有下位类——诊断、治疗和预防（用蓝色表示节点）；执行完整个语句时，才查询出图 8 内包含的所有节点，并且只有实行 RETURN 返回语句，才会将这些节点显示在图中（见图 8）。

该知识实体是站在中西医角度分别从不同方面对新冠展开研究，归纳总结出新冠的相关结论，不仅促进了人们对新冠的深入了解，而且有利于充分发挥医疗系统在诊断治疗中的作用，便于更好地制定防控措施及救援方案。因此次疫情的影响范围广泛，全球各地的科研人员对 COVID-19 的病毒来源、传播途径、疫情预测、诊断和治疗等各个方面展开研究，从陈莹 [19] 等的可视化研究中可以看出国内外对新冠的重视，相信在科研人员的努力下能够早日攻克难关。

5 基于知识实体的应急突发公共卫生事件数据平台构建

如前文所述，知识实体图是对物理世界的符号表达，描述现实事件中存在的一些概念以及它们之间的联系 [20]。应急资源平台基于知识实体图而构建，采用自底向上的运作模式。搭建基于知识实体的应急资源平台，以突发公共卫生事件为基础，挖掘当前新冠疫情内部以及突发事件之间的多元化关系，可视化地展现给用户，可以使得信息的获取变得方便迅速且高效。用户通过智能搜索和智能推荐能够快速精准地收集信息，更好地利用信息资源做好应急防控措施。

如图 9 所示，应急资源平台包含数据层、知识库构建、数据存储和数据应用层四部分。

第一部分数据层位于平台的最底层，主要包括数据采集、数据清洗、数据转换及标准化，经过数据处理后得到的非结构化数据和半结构化数据作为构建知识实体库的数据源。在该层次掌握国内外有关突发公共卫生事件的研究现状，进行逻辑推理，从大数据中分析突发公共卫生事件的规律，让事物的发展变得有迹可循，从而形成搭建知识平台底层的架构。

原始数据按照数据的结构化程度可以划分为结构化数据、半结构化数据和非结构化数据。结构化数据即行数据，是存储在数据库里，可以用二维表结构来逻辑表达实现的数据，主要通过关系型数据库进行存储和管理。半结构化数据，是结构化数据的一种形式，虽不符合关系

知识管理论坛，2020(3):175-190
DOI：10.13266/j.issn.2095-5472.2020.016

型数据库或其他数据表的形式关联起来的数据模型结构，但包含相关标记，用来分隔语义元素以及对记录和字段进行分层。非结构化数据是数据结构不规则或不完整，没有预定义的数据模型，不方便用数据库二维逻辑表来表现的数据。包括关于某一类应急突发事件所有格式的办公文档、文本、图片、HTML、各类报表、图像和音频、视频信息等。

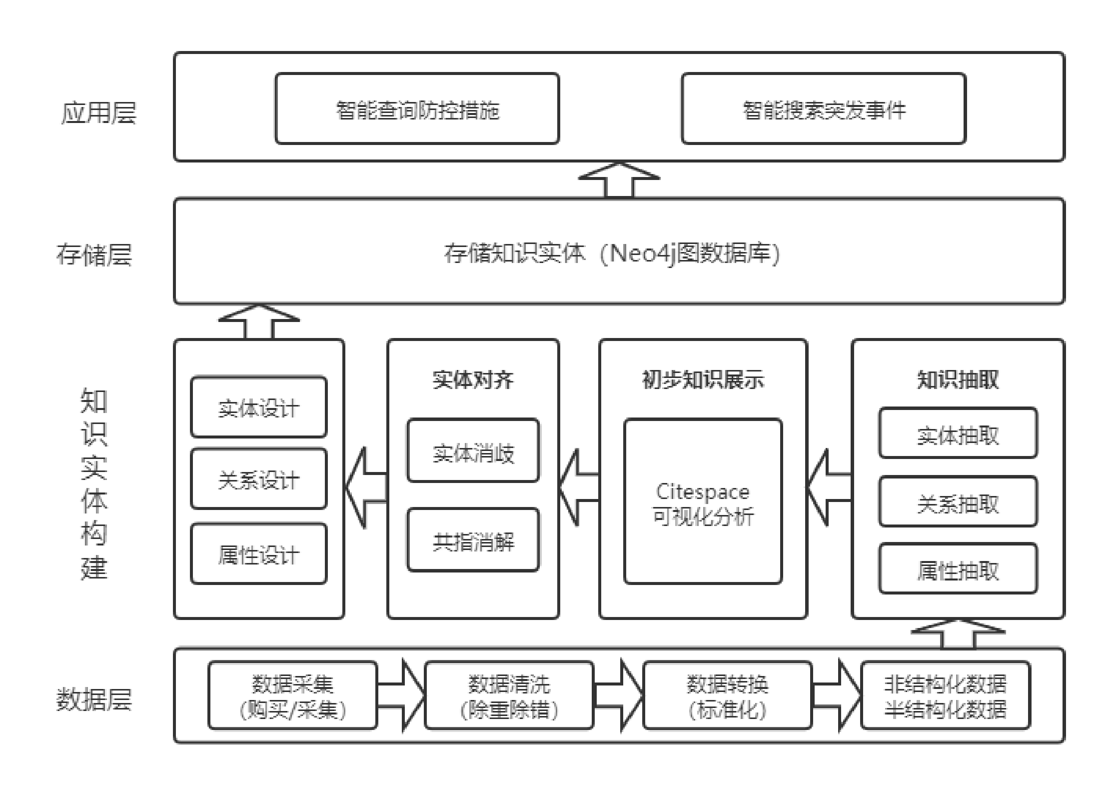


图 9 突发公共卫生事件数据平台总体框架

第二部分为知识实体构建，位于数据层的上方，包括知识抽取、实体对齐和知识设计，是平台的重要组成，有着承上启下的作用。通常一个知识图谱的构成包括实体集 (实体的个数)、关系集 (关系的数目) 以及对应的三元组。本体构建模块中主要是对突发公共卫生事件概念表达，用抽象的概念表达真实的存在，是一个迭代的过程。借助本体的约束来规范实体、关系以及实体属性之间的联系，可以有效解决“信息孤岛”的现象，形成本体化的知识表达。

本文提取了突发公共卫生事件的实体和关系，用实体、属性、属性值这样的三元组来表达事实。首先搭建出“突发公共卫生事件”和“public health emergencies”两个一级知识实体，再根据中国知网和 WOS 文献中关键词、作者和机构聚类的数据，划分出下一级的知识实体，完成了知识实体的构建。之后分析突发公共卫生事件的节点属性，通过属性值刻画出实体之间的内在特性，用关系来连接两个实体，深入挖掘内部联系。

在该过程中，知识抽取涉及的关键技术包括实体抽取、关系抽取和属性抽取。本文从原始的非结构化和半结构化数据中，即从国内的中国知网和国外的 WOS 检索到的关于突发公共卫生事件的相关文献中，提取关键词、作者和机构等关键知识实体，把排名前十的文献作者和机构进行梳理统计，将汇聚的实体知识加以规范，完成本体模块模型构建，利用 Citespace

可视化软件进行初步知识展示。获取到实体、关系以及实体的属性信息后，接下来要将这些碎片化的信息进行融合，进行知识实体消歧、共指消解，它的过程有实体链接，适用于半结构化和非结构化数据通过信息抽取提取出来的数据；知识合并，主要对结构化数据(如外部知识库和关系数据库)进行处理。

第三部分为数据存储层，位于知识实体库的上层，其构建基于知识实体库的实体数据。在知识实体有大量数据的复杂的情况下，就会显得传统的关系数据存储方式有些低效耗时，因此使用图数据库在查询工作中会明显提升工作效率，更易于图查询和搜索。

在该层次提取新冠的知识实体，从中医药学和西医药学两个方面划分知识实体，归纳新冠的相关结论。把最底层的数据转化为具有决策价值的信息，将有价值的信息沉淀下来与人的知识体系融合，形成一个完整的知识体系。用图的形式存储，选择图数据库 Neo4j 作为描述实体属性与实体之间关系的存储介质。在 Neo4j 图数据库中建立节点、关系、属性等，形成知识图谱，就可以进行下一步的查询了。

将数据存储在网络上，基于图的搜索，具有完全事务管理功能，可以很好地支撑动态数据特性的应用需求。根据本体模块对应的相关概念建立于实体数据之间的关系，实现概念的实体、关系、属性的知识实体可视化。

第四部分数据应用层位于平台的顶端，是平台的最终形式，主要面向对象有普通大众、政府工作人员、医务工作者、科研人员、教育教学者等，是一种基于知识为用户提供服务的智慧搜索引擎，能够结合生活实际，通过一般规律找到解决问题的方法，实现从了解问题到明白如何解决问题的转变。智能查询应急防控和智能搜索突发事件等功能的实现，便于用户更加精准有效地得到自己需要的知识，为应急救援决策提供切实可行的应对策略。在大数据的支撑下，利用知识实体图实现搜索结果的可视化、体系化，使知识体系构成网状节点，展现更加精准的信息。应急事件大数据平台不是简单的数据输入和输出，而是一个以信息资源为基础，对来自不同领域的信息进行处理，通过内部实体的筛选与融合构建的面向大众的共享资源平台。

6 结语与思考

应急突发事件下，信息资源的高效利用对事件的处理影响是巨大的，而资源的不充分利用会成为有限资源的一种浪费。利用大数据实现从简单信息服务到知识服务的提升是必然的，所以利用大数据技术挖掘、分析有用的数据，对资源进行充分有效的利用变得尤为重要。本文以疫情防控和突发公共卫生事件为基础，研究经历了以下阶段：①搜集相关资料，建立基于知识实体的突发卫生公共事件平台搭建思路；②在中国知网和 WOS 数据库，分别以“突发公共卫生事件”“public health emergencies”关键词进行高级检索；③通过构建知识实体，深入主题进行阐述，挖掘当前新冠疫情内部以及突发事件之间的多元化关系；④对基于知识实体的应急大数据的构建进行细化，将问题延伸。在此基础上构建应急资源平台，发挥信息化对辅助疫情诊断、创新医疗模式、提升服务效率等方面的作用，从而做好疫情发现、防控和应急处置工作。

基于知识图谱构建应急资源平台，在理论层面，将知识实体的理论方法应用于智库的建设，且此知识实体是不断扩充和生长的，不仅丰富智慧数据平台建设的知识实体方法理论，也有效地解决了数据平台的生命力和持续性的问题；在应用层面，依靠权威的文献数据库平台来获取知识实体，可以使信息获取变得方便迅速且高效，且对于多年来发生的一系列公共卫生突发事件进行实体提取，可以在同类型事件下查询到可借鉴的预防、治疗、应对等方法，为公共卫生事件应急决策提供了有力支撑。

现阶段知识实体的构建在我国还处于发展初期，许多技术及知识获取的算法还有待改善

知识管理论坛，2020(3):175–190
DOI：10.13266/j.issn.2095–5472.2020.016

和发展。然而要特别指出的是，突发公共卫生事件数据平台的构建是一个复杂的过程，其中会涉及到不同的环境背景、主体、文化以及运行机制等。本文构建的基于知识实体的应急事件大数据平台仅仅是一个模型，现有的研究数量有一定的局限，面对多变难掌握的环境因素，相关结论需要进一步验证与完善，理论与实践并重，人文与技术相融合。在下一步规划设想中，将会增加更多的数据集，构建基于知识实体的应急事件大数据平台，应用于不同领域，有效应对不同类型的突发事件。

参考文献：

[1] 习近平 . 全面提高依法防控依法治理能力 健全国家公共卫生应急管理体系 [J]. 当代广西 , 2020(5): 4-5.

[2] DUNLOP A L, LOGUE K M, BELTRAN G, et al. Role of academic institutions in community disaster response since September 11, 2001[J]. Disaster medicine and public health preparedness, 2011, 5(3): 218-226.

[3] VIELOT N A, HORNEY JENNIFER A. Can merging the roles of public health preparedness and emergency management increase the efficiency and effectiveness of emergency planning and response?[J]. International journal of environmental research and public health, 2014, 11(3): 2911-2921.

[4] HU JIAXIANG, CHEN CHAO, KUAI TINGTING. Improvement of emergency management mechanism of public health crisis in rural China: a review article[J]. Iranian journal of public health, 2018, 47(2): 156-165.

[5] ZHONG S, HE T, LI M, et al. An emergency resource scheduling model based on edge computing[M]// Artificial Intelligence for Communications and Networks. Cham: Springer, 2019.

[6] 柯丹倩 . 国内外突发事件驱动的应急决策情报研究现状及展望 [J]. 现代情报 , 2015, 35(8): 12-16.

[7] 张永领 , 陈璐 . 基于情景分析的应急资源保障能力评价模型研究 [J]. 中国安全科学学报 , 2014, 24(12): 151-157.

[8] 郭路生 , 刘春年 , 李瑞楠 . 面向公众服务的应急信息资源目录体系的构建研究 [J]. 图书馆学研究 , 2016(7): 41-49, 23.

[9] 屈腾佼 , 谷仕艳 , 李萌竹 , 等 . 中国卫生应急管理发展现状及面临挑战 [J]. 中国公共卫生管理 , 2019(4): 433-435.

[10] 李品 , 许林玉 , 杨建林 . 面向智库服务的情报研究 [J]. 情报学报 , 2020, 39(2): 135-147.

[11] 操玉杰 , 李纲 , 毛进 , 等 . 大数据环境下面向决策全流程的应急信息融合研究 [J]. 图书情报知识 , 2018(5): 95-104.

[12] 项灵辉 , 顾进广 , 吴钢 . 基于图数据库的 RDF 数据分布式存储 [J]. 计算机应用与软件 , 2014, 31(11): 35-39.

[13] 林启胜 , 王磊 , 周喜 , 等 . 基于图数据库的文献检索方法优化与实现 [J]. 微电子学与计算机 , 2017, 34(10): 63-67.

[14] 畜牧水产 . 国家突发公共卫生事件应急预案 [J. 中国食品卫生杂志 , 2006, 18(4): 366-373.

[15] 林枫 . 分布式键值对存储系统的设计与实现 [D]. 广州 : 华南理工大学 , 2017.

[16] 张凤军 . 基于 Neo4j 图数据库的社交网络数据的研究与应用 [D]. 长沙 : 湖南大学 , 2016.

[17] 贾明 , 王虹 , 宋春利 , 等 . 新型冠状病毒肺炎文献整理及研究概述 [J]. 陕西医学杂志 , 2020, 49(3): 259-263, 266.

[18] 赵钢 , 曹莹莹 , 于文慧 , 等 . 基于中医经典理论探讨新型冠状病毒肺炎的防治 [J]. 江苏中医药 , 2020, 52(4): 38-42.

[19] 陈莹 , 郭怡博 , 郭然 , 等 . 基于文献计量学的新型冠状病毒肺炎（COVID-19）研究可视化分析 [J]. 中国中药杂志 , 2020, 45(10): 2239-2248.

[20] 刘峤 , 李杨 , 段宏 , 等 . 知识图谱构建技术综述 [J]. 计算机研究与发展 , 2016, 53(3): 582-600.

作者贡献说明：

冯　鑫：数据收集与分析，实验验证；
李　雪：理论构建，论文撰写；
闫　月：数据收集与分析，实验验证，论文撰写；
李佳培：结论分析，论文指导；
刘梦瑶：语言润色，论文修改与最终版本修订；
吴　晔：结论分析，论文指导。


知识管理论坛
2020 年 第 3 期(总第 27 期)

Study on Construction of Emergent Public Health Event Data Platform Based on Knowledge Entity

Feng Xin[1] Li Xue[2] Yan Yue[2] Li Jiapei[1] Liu Mengyao[1] Wu Ye[3,4]
[1]Management College, Hebei GEO University, Shijiazhuang 050031
[2]Huaxin College of Hebei Geo University, Shijiazhuang 050031
[3]Computational Communication Research Center, Beijing Normal University, Zhuhai 519087
[4]School of Journalism and Communication, Beijing Normal University, Beijing 100875

Abstract: [Purpose/significance] The public health emergencies endanger social public life health around global seriously. The realization of information resources for the intelligent storage, query, and knowledge of emergency organization and output has important reference significance for the integration sharing of scientific research data and knowledge management in emergency field. **[Method/process]** This paper selected the China Knowledge Network and Web of Science as the retrieval platform. Set health emergencies as the search conditions to crawl valid data, and used literature measurement software to extract knowledge data processing and fusion, and did cluster visualization analysis of the high frequency vocabulary, the author of the literature, and institutions. Then the knowledge entity, attribute and relation were stored in the Neo4j graph database, and the complex relation between knowledge entities was established, finally built the public health emergencies data platform model. **[Result/conclusion]** Emergency event wisdom platform is a knowledge data sharing resource platform facing the public which is based on the information resources. It can processes the information from different areas, and is built by the screening and fusion of internal entity. The platform can realize information resources flow by upper and lower interaction, to play the role of information resources of emergency management work.

Keywords: public health emergencies emergency service knowledge entity COVID-19